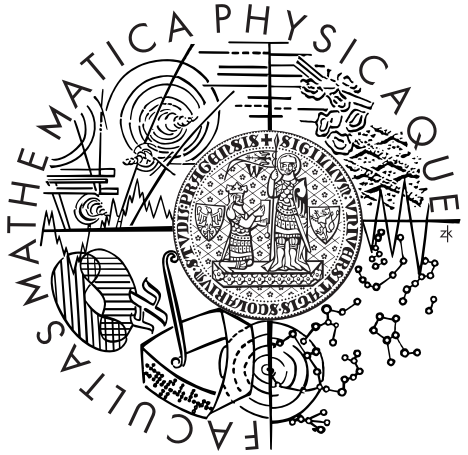
Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Jiří Sýkora

Kombinatorika hashovacích funkcí

Department of Algebra

Supervisor of the master thesis: doc. Mgr. Štěpán Holub, Ph.D.

Study programme: Mathematics

Specialization: Mathematical Methods of Information Security

Prague 2012

I would like to thank my supervisor, doc. Mgr. Štěpán Holub, Ph.D., for his patience and insightful remarks.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a licence agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, 13th April 2012 Jiří Sýkora

Název práce: Kombinatorika hashovacích funkcí

Autor: Jiří Sýkora

Katedra: Katedra algebry

Vedoucí diplomové práce: doc. Mgr. Štěpán Holub, Ph.D., Katedra algebry

Abstrakt: V této práci se zabýváme hašovacími funkcemi. Soustředíme se především na známou Merkle-Damgårdovu konstrukci a její zobecnění. Ukazujeme, že ani tato zobecněná kontrukce není odolná proti útokům hledajícím multikolize. Zásadní roli při tvorbě našeho útoku hraje kombinatorika na slovech. Ukazuje se totiž, že v dostatečně dlouhých slovech s omezeným počtem výskytů jednotlivých symbolů se nutně musí objevovat určité pravidelnosti. V této oblasti předvádíme vlastní původní výsledky, kterými zlepšujeme dříve publikované odhady, čímž snižujeme složitost útoku. Z toho plyne, že zobecněné hašovací funkce jsou zajímavé spíše z teoretického než praktického hlediska.

Klíčová slova: hašovací funkce, multikolize, kombinatorika na slovech

Title: Combinatorics of hash functions

Author: Jiří Sýkora

Department: Department of Algebra

Supervisor: doc. Mgr. Štěpán Holub, Ph.D., Department of Algebra

Abstract: In this thesis, we study hash functions. We focus mainly on the famous Merkle-Damgård construction and its generalisation. We show that even this generalised construction is not resistant to multicollision attacks. Combinatorics on words plays a fundamental role in the construction of our attack. We prove that regularities unavoidably appear in long words with bounded number of symbol occurences. We present our original results concerning regularities in long words. We lower some earlier published estimates, thus reducing the comlexity of the attack. Our results show that generalised iterated hash functions are interesting rather from the theoretical than practical point of view.

Keywords: hash functions, multicollisions, combinatorics on words

Contents

Introduction

Hash functions play an important role in cryptography. They have applications in digital signatures, password verification and other forms of authentication. Therefore, it is important that they have all the desired properties and cannot be easily attacked.

A (cryptographic) *hash function* is a function mapping binary strings of arbitrary length to fixed-length binary strings called *hash values*. The input string is often called a *message*. For practical purposes, we want the function to be computationally efficient. Furthemore, it is desirable that a hash function behaves randomly. That means that a small change in the input message results in a big change in the corresponding hash value. It should be also difficult to find two messages with the same hash value, or, when given a message, to find a different message with the same hash value.

Many hash functions, including some of the widely used ones as SHA-1 or MD5, are built in the form of the Merkle-Damgård construction [8, 1]. It means that the input message is divided into blocks of the same length, and these blocks are successively processed by a compression function. In this thesis we also call these hash fuctions *iterated hash functions*. However, many of iterated hash functions have been "broken"; there have been found attacks on them with complexity lower than we would expect of an ideal hash function. Naturally, the question arises as to whether it is possible to modify the Merkle-Damgård construction to prevent these attacks. We could, for example, compress the blocks of the input message in different order or use some of them more then once. Then we need to know whether such a generalisation improves the security significantly.

In this thesis, we study generalised iterated hash functions. Our work is based on [5] and [6], where the authors described a multicollision attack on generalised iterated hash fuctions. In the following chapter, we introduce basic terminology of words and languages as well as basics of partial orders on words. We also describe the birthday paradox. In the second chapter, we state the formal definition of a generalised iterated hash function. We also mention some earlier results concerning hash functions and collisions. Chapter 3 contains our main results; we prove that certain regularities in long words are unavoidable, and we reduce the bound from [6]. Finally, in the fourth chapter, we describe the relation between the combinatorial results and a multicollision attack on so-called q-bounded generalised iterated hash functions, thus proving that they are not secure.

1. Terminology and basics

In this chapter, we introduce basics of words, languages and partial orders. As will be shown later, they are closely related to hash functions. For the sake of convenience, we use the same terminology as in [5]. We also describe the birthday paradox which is used quite often when studying hash functions.

1.1 Words and languages

We denote the set of all *positive integers* by $\mathbb{N}_+$ and the set of all *natural numbers* by $\mathbb{N}$, i.e. $\mathbb{N} = \mathbb{N}_+ \cup \{0\}$. For each $l \in \mathbb{N}_+$ we define $\mathbb{N}_l = \{1, 2, \ldots, l\}$.

An *alphabet* is a finite, nonempty set of elements called *symbols* or *letters*. A *word* over an alphabet is a finite sequence of symbols from the alphabet. Therefore, assuming we have a word w over an alphabet A, we can write $w = x_1 x_2 \cdots x_n$, where n is a nonnegative integer and $x_i \in A$ for $i \in \{1, 2, \ldots, n\}$. The integer n is called the *length* of w and denoted $|w|$. If n is equal to zero, then we call w the *empty word*, which is often denoted by ϵ. Let A^* (resp. A^+) be the set of all words (resp. all nonempty words) over A. Similarly, we define a^* (resp. a^+) as the set of all words (resp. all nonempty words) over the alphabet $\{a\}$. By $|w|_a$ we understand the number of occurences of the letter a in w. The *alphabet* of w is defined as $\mathrm{alph}(w) = \{a \in A \,\big|\, |w|_a > 0\}$. Let $a \in \mathrm{alph}(w)$ and $i \in \{1, 2, \ldots, |w|_a\}$. Then we denote the i-th occurence of the symbol a in the word w by $a^{w,i}$. We say that $v = v_1 v_2 \cdots v_m$, where $v_1, v_2, \ldots, v_m \in A^*$, is a *subword* of w if $w = x_0 v_0 x_1 \cdots v_m x_m$ for some $x_0, x_1, \ldots, x_m \in A^*$. We call a subword v of w a *factor* of w if $w = x_0 v x_1$, where $x_0, x_1 \in A^*$. We say that words $w_1, w_2, \ldots, w_m$ form a *factorisation* of the word w if $w = w_1 w_2 \cdots w_m$.

Let A and B be alphabets. A mapping $h \colon A^* \to B^*$ is a *morphism* if $h(uv) = h(u)h(v)$ for all $u, v \in A^*$. If $B \subseteq A$, then the *projection morphism* from A^* into B^*, denoted by π_B^A (or π_B, when A is clear from the context), is defined by $\pi_B^A(b) = b$ for each $b \in B$ and $\pi_B^A(a) = \epsilon$ for each $A \setminus B$. We also define $(w)_B = \epsilon$ if $\pi_B(w) = \epsilon$ and $(w)_B = a_1 a_2 \cdots a_s$ if $\pi_B(w) = a_1^+ a_2^+ \cdots a_s^+$, where $s \in \mathbb{N}_+$, $a_1, a_2, \ldots, a_s \in B$ and $a_i \neq a_{i+1}$ for $i = 1, 2, \ldots, s-1$.

A word w over an alphabet A is a *permutation* of A if $|w|_a = 1$ for each $a \in A$. A *language* over the alphabet A is an arbitrary subset $L \subseteq A^*$.

1.2 Partial orders on words

Let $(X, \prec)$ be a partially ordered set. The elements $x, y \in X$, $x \neq y$ are *incomparable* if neither $x \prec y$ nor $y \prec x$ holds. The elements $x_1, x_2, \ldots, x_n \in X$ form a *chain* in $(X, \prec)$ if $x_i \prec x_{i+1}$ for each $i \in \{1, 2, \ldots, n-1\}$. The *length* of a chain is equal to the number of elements forming the chain. We denote the length of a chain c by $|c|$. A *chain decomposition* of $(X, \prec)$ is a set of chains $\{c_i\}_{i \in I}$ such that $\{C_i\}_{i \in I}$ is a partition of X, where $C_i = \{x \in X \,|\, x \text{ occurs in } c_i\}$. Now suppose that X is finite. We call the cardinality of the largest set $Y \subseteq X$ consisting of pairwise incomparable elements the *maximum number of incomparable elements* (of $(X, \prec)$). The *minimum chain decomposition size* of $(X, \prec)$ is the smallest

positive integer m such that there exist chains $c_1, c_2, \ldots, c_m$ in $(X, \prec)$ for which $\{c_i\}_{i=1}^m$ is a chain decomposition of $(X, \prec)$. We also define the *maximum chain length* of $(X, \prec)$ as the greatest positive integer m such that there exists a chain of legth m in $(X, \prec)$.

Now we can introduce a partial order on symbols in a word. Let α be a nonempty word. We define the binary relation $\prec_\alpha$ on $\mathrm{alph}(\alpha)$ as follows. For each $a, b \in \mathrm{alph}(\alpha)$, $a \prec_\alpha b$ holds if and only if $a \neq b$ and all occurences of a in α lie before the first occurence of b in α. Quite obviously, $\prec_\alpha$ forms a partial order on $\mathrm{alph}(\alpha)$. We call some elements *independent* (with respect to $\prec_\alpha$) if they form a chain in $(\mathrm{alph}(\alpha), \prec_\alpha)$.

The following famous theorem of Dilworth [2] gives a connection between the number of incomparable elements and the minimum chain decomposition size.

Theorem 1. *Let $(X, \prec)$ be a finite partially ordered set. Then the maximum number of incomparable elements of $(X, \prec)$ is equal to the minimum chain decomposition size of $(X, \prec)$.*

We can apply Dilworth´s theorem on words and derive the following useful lemma (see Lemma 4.8. in [5]). We repeat the proof for the sake of completeness.

Lemma 1. *Let m and n be positive integers and α a word such that $\mathrm{alph}(\alpha) \geq m \cdot n$. Then either (i) the maximum chain length of $(\mathrm{alph}(\alpha), \prec_\alpha)$ is at least m; or (ii) the maximum number of pairwise incomparable elements in $(\mathrm{alph}(\alpha), \prec_\alpha)$ is greater than n.*

Proof. Suppose that the maximum chain length in $(\mathrm{alph}(\alpha), \prec_\alpha)$, denoted by d, is less than m. Let t be the minimum number of chains needed to cover $(\mathrm{alph}(\alpha), \prec_\alpha)$. Obviously

$$m \cdot n \leq |\mathrm{alph}(\alpha)| \leq d \cdot t.$$

Since $d < m$, we have $t > n$. By Dilworth´s theorem the maximum number of pairwise incomparable elements of $(\mathrm{alph}(\alpha), \prec_\alpha)$ is equal to t. □

1.3 Birthday paradox

The birthday paradox or also birthday problem is a basic concept needed to understand collisions in hash functions. Suppose we have a set of n elements. We pick k elements uniformly at random (we can pick one element more times). What is the probability p that we pick one element at least twice? Another formulation of this problem could be as follows. What is the probability that among k people there are at least two with the same birthday? In this formulation, we have $n = 365$. We shall discuss the general case. Obviously, the probability is equal to one for $k > n$. Thus, we can restrict ourselves to the case $k \leq n$. Firstly, we calculate the probality $\bar{p}$ that all our picks are different. We have

$$\bar{p} = 1 \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-(k-1)}{n} = 1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right).$$

Since $1 + x \leq e^x$ for each real number x, we get

$$\bar{p} \leq e^0 \cdot e^{-\frac{1}{n}} \cdots e^{-\frac{k-1}{n}} = e^{-\frac{1}{n}(1+2+\cdots+k-1)} = e^{-\frac{k(k-1)}{2n}}.$$

Our probability p is equal to $p = 1 - \bar{p}$. Say we want p to be at least 0.5. In other words, we need $\bar{p} \leq 0.5$. We calculate as follows:

$$\begin{aligned} \bar{p} &\leq \frac{1}{2} \\ e^{-\frac{k(k-1)}{2n}} &\leq \frac{1}{2} \\ -\frac{k(k-1)}{2n} &\leq -\log 2 \\ k(k-1) &\geq 2n \log 2 \\ k^2 - k - 2n \log 2 &\geq 0. \end{aligned}$$

By solving the quadratic inequality we get that $k \geq \frac{1}{2} + \sqrt{\frac{1}{4} + 2n \log 2}$ implies $p \geq \frac{1}{2}$.

Therefore, we need to pick only about $\sqrt{n}$ elements to get at least one twice. Intuitively, most of the people would expect that k needs to be about $\frac{n}{2}$. Applying our result to the case $n = 365$, the lower bound on k is approximately 23. Thus, among 23 people there are at least two sharing their birthday with the probability over 0.5. This is the reason why this problem is called the birthday paradox.

2. Hash functions

In this chapter, we describe hash functions and their generalisations using a notation consistent with [5]. Some earlier ideas on hash functions and multicollisions can be also found in [3, 4, 9].

2.1 Basic definitions and notation

When working with iterated hash functions, we suppose that the original message is given in a block representation form, i.e. the message is divided, and if necessary padded, into blocks of equal size. We also assume that all our messages are over the binary alphabet $\{0,1\}$.

Definition 1. A *hash function of length* n (where $n \in \mathbb{N}_+$) is a mapping $f : \{0,1\}^* \to \{0,1\}^n$.

By a *preimage* of a given hash value y we understand $x \in \{0,1\}^*$ such that $f(x) = y$. Let x be a given messsage. A *second preimage* of $y = f(x)$ is $x' \in \{0,1\}^*$, $x' \neq x$ such that $f(x') = y$. Let k be a positive integer. A *k-collision* in f is a set $A \subseteq \{0,1\}^*$ satisfying $|A| = k$ and $f(x) = f(y)$ for all $x, y \in A$. We often call a 2-collision simply a collision in f.

An ideal hash function is a *variable input length random oracle*, which means that for each $x \in \{0,1\}^*$, the value $f(x) \in \{0,1\}^n$ is chosen uniformly at random. For an ideal hash function f, we need, on average, to hash $O(2^n)$ messages to find a preimage or a second preimage. To find a k-collision with a probability approximately 0.5 we need to hash

$$(k!)^{\frac{1}{k}} 2^{\frac{n(k-1)}{k}}$$

messages. This follows from the generalised birthday paradox, for details see [10].

Hash functions are often constructed iteratively. In the following, we describe an iterated hash function and its generalisation.

Definition 2. A *compression function* (of block size m and lenght n) is a mapping $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^n$, where $m > n$ are positive integers.

An ideal compression function is also a random oracle, in this case a so-called *fixed input length random oracle*.

For a compression function $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^n$, we define the function $f^+ : \{0,1\}^n \times (\{0,1\}^m)^+ \to \{0,1\}^n$ inductively as follows. Let $h \in \{0,1\}^n, y_1 \in \{0,1\}^m$, and $y_2 \in (\{0,1\}^m)^+$. Then $f^+(h, y_1) = f(h, y_1)$ and $f^+(h, y_1 y_2) = f^+(f(h, y_1), y_2)$. Note that $f^+(h, y\, y') = f^+(f^+(h, y), y')$ for all $y, y' \in (\{0,1\}^m)^+$.

Let u be a word in $(\{0,1\}^m)^+$. Then we can write $u = u_1 u_2 \cdots u_l$, where $l \in \mathbb{N}_+$ and $u_i \in \{0,1\}^m$ for each $i \in \{1, 2, \ldots, l\}$. Define the morphism $\bar{u} : \mathbb{N}_l^* \to \{0,1\}^*$ by $\bar{u}(i) = u_i$ for each $i \in \mathbb{N}_l$. Suppose $\alpha \in \mathbb{N}_l^+$ is given. It can be expressed as $\alpha = i_1 i_2 \cdots i_s$, where $s \in \mathbb{N}_+$ and $i_j \in \mathbb{N}_l$ for $j = 1, 2, \ldots, s$. Then $\bar{u}(\alpha) = u_{i_1} u_{i_2} \cdots u_{i_s}$. Obviously $\bar{u}(\alpha)$ is a word containing blocks from u in the order and multiple determined by α.

Now we can define the *iterated compression function* $f_\alpha : \{0,1\}^n \times \{0,1\}^{ml} \to \{0,1\}^n$ based on α and f by $f_\alpha(h,u) = f^+(h,\bar{u}(\alpha))$ for each $h \in \{0,1\}^n$ and $u \in \{0,1\}^{ml}$. It is quite straightforward to see that for $\alpha = \alpha_1\alpha_2$, where $\alpha_1, \alpha_2 \in \mathbb{N}_l^+$, and for all $h \in \{0,1\}^n$ and $u \in \{0,1\}^{ml}$ the following equalities hold:

$$\begin{aligned} f_\alpha(h,u) &= f^+(h,\bar{u}(\alpha)) = f^+(h,\bar{u}(\alpha_1)\bar{u}(\alpha_2)) = \\ &= f^+(f^+(h,\bar{u}(\alpha_1)),\bar{u}(\alpha_2)) = f_{\alpha_2}(f_{\alpha_1}(h,u),u). \end{aligned}$$

We say that a set $A \subseteq \{0,1\}^{ml}$ is a *k-collision* in f_α with the initial value h_0, where $k \in \mathbb{N}_+$ and $h_0 \in \{0,1\}^m$, if $|A| = k$ and $f_\alpha(h_0,u) = f_\alpha(h_0,v)$ for all $u,v \in A$. The k-collision A is called *nontrivial* if, for each $u = u_1u_2\cdots u_l$ and $v = v_1v_2\cdots v_l$ in A such that $u_i, v_i \in \{0,1\}^m$ for $i = 1,2,\ldots,l$, the equality $u_j = v_j$ holds for each $j \in \mathbb{N}_l \setminus \mathrm{alph}(\alpha)$.

Definition 3. Let $\alpha_j \in \mathbb{N}_j^+$ and $\mathrm{alph}(\alpha_j) = \mathbb{N}_j$ for each $j \in \mathbb{N}_+$. Denote $\hat{\alpha} = (\alpha_1, \alpha_2, \ldots)$. Then we define the *generalised iterated hash function*

$$H_{\hat{\alpha},f} : \{0,1\}^n \times (\{0,1\}^m)^+ \to \{0,1\}^n$$

based on $\hat{\alpha}$ and f as follows: for the initial value $h_0 \in \{0,1\}^n$ and the message x consisting of j blocks (of the lentgh m) let $H_{\hat{\alpha},f}(h_0,x) = f_{\alpha_j}(h_0,x)$.

Remark 1. A traditional iterated hash function $H : (\{0,1\}^m)^+ \to \{0,1\}^n$ based on f (with the initial value $h_0 \in \{0,1\}^n$) is defined by $H(u) = f^+(h_0,u)$ for all $u \in (\{0,1\}^m)^+$. This definition is equivalent to a generalised iterated hash function $H_{\hat{\alpha},f} : \{0,1\}^n \times (\{0,1\}^m)^+ \to \{0,1\}^n$ where $\hat{\alpha} = (1, 1\cdot 2, 1\cdot 2\cdot 3, \ldots)$ and the initial value is fixed as h_0.

We say that a set $A \subseteq (\{0,1\}^m)^+$ is a *k-collision* in $H_{\hat{\alpha},f}$ with the initial value h_0, where $k \in \mathbb{N}_+$ and $h_0 \in \{0,1\}^m$, if $|A| = k$ and for all $u,v \in A$, $|u| = |v|$ and $H_{\hat{\alpha},f}(h_0,u) = H_{\hat{\alpha},f}(h_0,v)$.

When constructing our attack, we assume that the attacker knows $\hat{\alpha}$ and thus knows in which order $H_{\hat{\alpha},f}$ compresses the blocks. On the other hand, the compression function f is a black box, i.e. the attacker does not know its internal structure and can only make queries on f and get responses.

2.2 Earlier results

In [4] Antoine Joux shows a multicollision attack on traditional iterated hash fuctions. For each $k \in \mathbb{N}_+$ he can find a 2^k-collision with the complexity $O(k\cdot 2^{\frac{n}{2}})$. The idea of the attack is quite simple yet efficient. Let $H : (\{0,1\}^m)^+ \to \{0,1\}^n$ be the given traditional iterated hash function based on the compression function f and h_0 be the initial value. It follows from the birthday paradox that, by hashing approximately $2^{\frac{n}{2}}$ messages, we can find a message set $\{u_{11}, u_{12}\}$ such that $f(h_0,u_{11}) = f(h_0,u_{12}) = h_1$. Similarly we find $\{u_{21}, u_{22}\}$ such that $f(h_1,u_{21}) = f(h_1,u_{22}) = h_2$. We can continue this process until we find $\{u_{k1}, u_{k2}\}$ such that $f(h_{k-1},u_{k1}) = f(h_{k-1},u_{k2}) = h_k$. The total number of queries on f is $O(k\cdot 2^{\frac{n}{2}})$. It is obvious that the set $\{u_{1i_1}u_{1i_2}\cdots u_{ki_k} \mid i_j = 1,2,\ j = 1,2,\ldots,k\}$ is a 2^k-collision in H.

In [9] Nandi and Stinson present a 2^k-collision attack on an iterated compression function f_α, where $f : \{0,1\}^n \times (\{0,1\}^m)^+ \to \{0,1\}^n$ and α satisfies the following conditions: (i) $\mathrm{alph}(\alpha)$ is large enough and (ii) $|\alpha|_a \leq 2$ for each $a \in \mathrm{alph}(\alpha)$. The complexity of their attack is

$$O(k^2 \cdot \ln r \cdot (n + \ln(\ln 2r)) \cdot 2^{\frac{n}{2}}).$$

In [3] Hoch and Shamir study generalised iterated hash functions. They show that for a function $H_{\hat{\alpha},f}$, where $|\alpha_i|_j \leq q$, $i \in \mathbb{N}_+$, $j \in \mathbb{N}_i$ and q is a fixed positive integer, there exists a multicollision attack. For a given $k \in \mathbb{N}_+$, a 2^k-collision in $H_{\hat{\alpha},f}$ can be constructed with the complexity $O(p(n,k)2^{\frac{n}{2}})$, where $p(n,k)$ is a polynomial. Similar results are obtained in [5]. In [6] the authors study regularities in long words with bounded number of symbol occurences which leads to reducing the complexity of the attack from [5]. In this thesis, we continue with their study and reduce the complexity even more.

3. Main combinatorial results

3.1 Regularities in long words

In this chapter, we focus on combinatorial properties of long words. It follows from the definition of generalised iterated hash functions that finding some regularities in long words can lead to a (multi-)collision attack. This attack is thoroughly explained in [5] and is also described in the following chapter.

Definition 4. Let m and q be positive integers. Then we define $N(m,q)$ as the smallest integer such that for every word α, where $|\mathrm{alph}(\alpha)| \geq N(m,q)$ and $|\alpha|_a \leq q$ for each $a \in \mathrm{alph}(\alpha)$, there exists $A \subseteq \mathrm{alph}(\alpha)$ which satisfies: $|A| \geq m$ and there exists $p \in \{1,2,\ldots,q\}$ and words $\alpha_1, \alpha_2, \ldots, \alpha_p$ such that $\alpha = \alpha_1\alpha_2\cdots\alpha_p$ and the word $(\alpha_i)_A$ is a permutation of A for all $i \in \{1,2,\ldots,p\}$.

The existence of such an integer for every m and q was proved in [6]. It was shown there that $N(m,q+1) \leq N(m^2-m+1,q)$ for all positive integers m and q. Obviously $N(m,1) = m$. Therefore $N(m,q) \leq m^{2^{q-1}}$.

The main goal of this thesis is to investigate the optimality of the upper bound on $N(m,q)$ presented in [6] and to improve the bound in cases where it is not perfect. Obviously $N(1,q) = 1$ for every positive integer q, which is exactly the value given in [6]. The following remark (see Remark 1. in [6]) treats the case $q = 2$.

Remark 2. Let $m \in \mathbb{N}_+$ and let

$$B = \{a_{i,j} | i = 1,2,\ldots,m-1,\ j = 1,2,\ldots,m\}$$

be an alphabet of $m(m-1)$ symbols. Let furthemore

$$\gamma_i = a_{i,1}a_{i,2}\cdots a_{i,m-1}a_{i,m}a_{i,m-1}a_{i,m-2}\cdots a_{i,1}$$

for $i = 1,2,\ldots,m-1$ and $\alpha = \gamma_1\gamma_2\cdots\gamma_{m-1}$. It is quite obvious that there does not exist an m-letter subalphabet $A \subseteq B$ such that either (i) $(\alpha)_A$ is a permutation of A or (ii) there exists a factorisation $\alpha = \alpha_1\alpha_2$ such that $(\alpha_1)_A$ and $(\alpha_2)_A$ are both permutations of A.

Therefore $N(m,2) = m^2 - m + 1$, so the bound in [6] is optimal also in the case $q = 2$.

Before we focus on the other cases, we introduce the following useful remark.

Remark 3. Suppose positive integers m and q are given, and we want to factorise a word α as in Definition 4. Thus we need to find a subalphabet $A \subseteq \mathrm{alph}(\alpha)$, $|A| \geq m$ and a factorisation of α into at most q factors α_i such that $(\alpha_i)_A$ is a permutation of A for each i. Let B be a subalphabet of $\mathrm{alph}(\alpha)$ and denote $\beta = \pi_B(\alpha)$. Suppose there exists $A' \subseteq B \subseteq \mathrm{alph}(\alpha)$ of cardinality at least m as well as words $\beta_1, \ldots, \beta_p$, $p \leq q$, such that $\beta = \beta_1\cdots\beta_p$ and for all $i \in \{1,2,\ldots,p\}$, the word $(\beta_i)_{A'}$ is a permutation of A'. Define $A = A'$. Then obviously, we can factorise the word α into $\alpha = \alpha_1\alpha_2\cdots\alpha_p$ so that β_i is a subword of α_i for all $i \in \{1,2,\ldots,p\}$. Evidently, the word α_i is a permutation of A for all $i \in \{1,2,\ldots,p\}$.

We shall now focus on the case $m = 2$.

Theorem 2. *Let q be a positive integer. Then $N(2, q) \leq q + 1$.*

Proof. Let α be a word satisfying $|\mathrm{alph}(\alpha)| \geq q + 1$ and $|\alpha|_a \leq q$ for each $a \in \mathrm{alph}(\alpha)$. Then there exist two symbols, denote them a_1 and a_2, such that $|\alpha|_{a_1} = |\alpha|_{a_2}$. Let $A = \{a_1, a_2\}$ and $\beta = \pi_A(\alpha)$. Without loss of generality we may assume that the word β starts with the symbol a_1. Therefore we have either

$$\beta = a_1^{\beta,1} \cdots a_1^{\beta,i_1} a_2^{\beta,1} \cdots a_2^{\beta,j_1} a_1^{\beta,i_1+1} \cdots a_1^{\beta,i_2} \cdots a_1^{\beta,i_k} a_2^{\beta,j_{k-1}+1} \cdots a_2^{\beta,j_k},$$

or

$$\beta = a_1^{\beta,1} \cdots a_1^{\beta,i_1} a_2^{\beta,1} \cdots a_2^{\beta,j_1} a_1^{\beta,i_1+1} \cdots a_1^{\beta,i_2} \cdots a_2^{\beta,j_{k-1}} a_1^{\beta,i_{k-1}+1} \cdots a_1^{\beta,i_k} \ .$$

In other words, β consists of blocks of the symbols a_1 and a_2 starting with a_1. Denote $i_0 = 0$, $j_0 = 0$. In the first case, we can define

$$\beta_n = a_1^{\beta,i_{n-1}+1} \cdots a_1^{\beta,i_n} a_2^{\beta,j_{n-1}+1} \cdots a_2^{\beta,j_n}$$

for all $n \in \{1, 2, \cdots, k\}$. Then, evidently, $\beta = \beta_1 \beta_2 \cdots \beta_k$, where $k \leq q$ since $|\beta|_{a_1} = |\beta|_{a_2} \leq q$. Moreover $(\beta_n)_A = a_1 a_2$ is a permutation of A for all $n \in \{1, 2, \ldots, k\}$ and we are done.

In the second case, we denote by l the greatest positive integer such that $j_{l-1} + 1 < j_l$. Such l must exist because $|\beta|_{a_1} = |\beta|_{a_2}$ and hence there is a block with at least two occurences of a_2. Then we can define

$$\beta_n = a_1^{\beta,i_{n-1}+1} \cdots a_1^{\beta,i_n} a_2^{\beta,j_{n-1}+1} \cdots a_2^{\beta,j_n} \text{ for } n \in \{1, \ldots, l-1\},$$

$$\beta_l = a_1^{\beta,i_{l-1}+1} \cdots a_1^{\beta,i_l} a_2^{\beta,j_{l-1}+1} \cdots a_2^{\beta,j_l-1},$$

$$\beta_{l+1} = a_2^{\beta,j_l} a_1^{\beta,i_l+1} \cdots a_1^{\beta,i_{l+1}},$$

$$\beta_n = a_2^{\beta,j_{n-2}+1} \cdots a_2^{\beta,j_{n-1}} a_1^{\beta,i_{n-1}+1} \cdots a_1^{\beta,i_n} \text{ for } n \in \{l+2, \ldots, k\}.$$

We get $\beta = \beta_1 \beta_2 \cdots \beta_k$, where $(\beta_n)_A = a_1 a_2$ for $n \in \{1, \ldots, l\}$ and $(\beta_n)_A = a_2 a_1$ for $n \in \{l+1, \ldots, k\}$. We can factorise the word β and therefore, according to Remark 3, we can also factorise the word α. □

Theorem 2 gives us an upper bound on $N(2, q)$. The following remark shows that the bound is sharp.

Remark 4. Let $\{a_1, a_2, \ldots, a_q\}$ be a set of q symbols and

$$\alpha = a_q a_{q-1} \cdots a_1 a_q a_{q-1} \cdots a_2 \cdots a_q a_{q-1} a_q \ .$$

Let $A = \{a_i, a_j\}$ be an arbitrary two-element subset of $\mathrm{alph}(\alpha)$ and suppose $i > j$. Then $\pi_A(\alpha) = a_i a_j a_i \cdots a_j a_i^+$ and it is quite straightforward to see that the word α cannot be factorised as in Definition 4. Therefore we get $N(2, q) = q + 1$.

The following theorem deals with the case $q = 3$.

Theorem 3. *Let m be a positive integer. Then $N(m, 3) \leq 2m^3 - m^2 + m$.*

Proof. Let α be a word such that $|\mathrm{alph}(\alpha)| \geq 2m^3 - m^2 + m$ and $|\alpha|_a \leq 3$ for each $a \in \mathrm{alph}(\alpha)$. Since $|\mathrm{alph}(\alpha)| \geq m \cdot (2m^2 - m + 1)$, Lemma 1 gives us these two possibilities:
(i) there exists a chain of length at least m in $(\mathrm{alph}(\alpha), \prec_\alpha)$. Then we can define A as the set of the elements of this chain. For this A, the word $(\alpha)_A$ is a permutation of A and we are done.
(ii) the maximum number of pairwise incomparable elements in $(\mathrm{alph}(\alpha), \prec_\alpha)$ is greater than $2m^2-m+1$. Denote by B the maximum set of pairwise incomparable symbols. We get $|B| \geq 2m^2 - m + 2$. We define $B_i = \{a \in B \mid |\alpha|_a = i\}$ for $i \in \{1, 2, 3\}$. Then evidently $|B_1| \leq 1$, otherwise there would be two symbols in B_1, denote them a and b such that $a \prec_\alpha b$ or $b \prec_\alpha a$, but we suppose that all symbols in B are pairwise incomparable.

Suppose now that $|B_2| \geq m$. Let $\beta = \pi_{B_2}(\alpha)$ and factorise the word $\beta = \beta_1\beta_2$ in the following way: we define β_1 as the shortest word containing all symbols from B_2 at least once. In other words, $\beta_1 = b_1 \cdots b_k$, $|\beta_1|_{b_k} = 1$ and $|\beta_1|_b \geq 1$ for all $b \in B_2$. Then $|\beta_1|_b = 1$ for all $b \in B_2$, otherwise we would get a symbol b_j such that $|\beta_1|_{b_j} = 2$ and therefore $b_j \prec_\beta b_k$, which yields $b_j \prec_\alpha b_k$, a clear contradiction with supposed incomparability of b_j and b_k. Hence we may define $A = B_2$ and the words $(\beta_1)_A$ and $(\beta_2)_A$ are permutations of A. It follows from Remark 3 that we can factorise the word α and we are finished.

The remaining case is $|B_2| \leq m - 1$ from which follows $|B_3| \geq 2m^2 - 2m + 2$, because $|B_1| \leq 1$. Denote $\gamma = \pi_{B_3}(\alpha)$. We factorise the word γ in a similar way as the word β in the previous paragraph, i.e. $\gamma = \gamma_1\gamma_2$ where γ_1 is the shortest word containing all symbols from B_3 at least once. So $\gamma_1 = c_1 \cdots c_l$, $|\gamma_1|_{c_l} = 1$ and $|\gamma_1|_c \geq 1$ for all $c \in B_3$. From the incomparability of the symbols it follows that $|\gamma_1|_c \leq 2$ for all $c \in B_3$. Therefore, we have either $|\gamma_1|_c = 2, |\gamma_2|_c = 1$, or $|\gamma_1|_c = 1, |\gamma_2|_c = 2$ for all $c \in B_3$. Denote $X = \{c \in B_3 \mid |\gamma_1|_c = 2\}$ and $Y = \{c \in B_3 \mid |\gamma_2|_c = 2\}$. Let $z = \max\{|X|, |Y|\}$. Then $z \geq |B_3|/2 = m^2 - m + 1$. Suppose $z = |X|$, the case $z = |Y|$ can be treated analogically. Denote $\delta = \pi_X(\gamma)$. Then $\delta = \delta_1\delta_2$, where $\delta_1 = \pi_X(\gamma_1)$ and $\delta_2 = \pi_X(\gamma_2)$. Since $|\mathrm{alph}(\delta_1)| = |X| \geq m^2-m+1 = N(m, 2)$, there exists a set $A \subseteq \mathrm{alph}(\delta_1)$, $|A| \geq m$, positive integer $p \in \{1, 2\}$ and words $\delta_{1,1}, \ldots, \delta_{1,p}$ such that $\delta_1 = \delta_{1,1} \cdots \delta_{1,p}$ and $(\delta_{1,i})_A$ is a permutation of A for all $i \in \{1, \ldots, p\}$. Clearly, $(\delta_2)_A$ is also a permutation of A and therefore we can factorise the word δ as $\delta = \delta_1\delta_2$, where $(\delta_1)_A, (\delta_2)_A$ are permutations of A, or $\delta = \delta_{1,1}\delta_{1,2}\delta_2$, where $(\delta_{1,1})_A, (\delta_{1,2})_A$ and $(\delta_2)_A$ are permutations of A. It follows from Remark 3 that we can also factorise the word α. □

For the general case, we have the following theorem.

Theorem 4. *Let m and q be positive integers, $q > 1$. Then*

$$N(m, q) \leq m \cdot \left(2 \cdot \left(\sum_{i=1}^{\lfloor \frac{q}{2} \rfloor} [N(N(m, i), q - i) - 1] \right) + N(m, q - 1) \right).$$

Proof. Let α be a word such that

$$|\mathrm{alph}(\alpha)| \geq m \cdot \left(2 \cdot \left(\sum_{i=1}^{\lfloor \frac{q}{2} \rfloor} [N(N(m, i), q - i) - 1] \right) + N(m, q - 1) \right)$$

and $|\alpha|_a \leq q$ for each $a \in \mathrm{alph}(\alpha)$. Lemma 1 gives us these two possibilities:
(i) there exists a chain of length at least m in $(\mathrm{alph}(\alpha), \prec_\alpha)$. Then we can define A as the set of the elements of this chain. For this A, the word $(\alpha)_A$ is a permutation of A and we are done.
(ii) there are at least

$$2 \cdot \left(\sum_{i=1}^{\lfloor \frac{q}{2} \rfloor} [N(N(m,i), q-i) - 1] \right) + N(m, q-1) + 1$$

pairwise incomparable elements in $(\mathrm{alph}(\alpha), \prec_\alpha)$. Denote by B the maximum set of pairwise incomparable symbols and define $B_{q-1} = \{a \in B \mid |\alpha|_a \leq q-1\}$, $B_q = \{a \in B \mid |\alpha|_a = q\}$. If $|B_{q-1}| \geq N(m, q-1)$ there exists $A \subseteq \mathrm{alph}(\beta)$, where β is defined as $\beta = \pi_{B_{q-1}}(\alpha)$, such that $|A| \geq m$ and we can factorise the word β as in Definition 4. Then, according to Remark 3, we can also factorise the word α.

If $|B_{q-1}| < N(m, q-1)$ we have

$$B_q \geq 2 \cdot \left(\sum_{i=1}^{\lfloor \frac{q}{2} \rfloor} [N(N(m,i), q-i) - 1] + 1 \right).$$

Denote $\gamma = \pi_{B_q}(\alpha)$. We factorise the word γ into $\gamma = \gamma_1\gamma_2$ where γ_1 is the shortest word containing all symbols from B_q at least once. So $\gamma_1 = c_1 \cdots c_l$, $|\gamma_1|_{c_l} = 1$ and $|\gamma_1|_c \geq 1$ for all $c \in B_q$. It follows from the incomparability of the symbols that $|\gamma_1|_c \leq q-1$ for all $c \in B_q$. Denote $X = \{c \in B_q \mid |\gamma_1|_c \geq q/2\}$ and $Y = \{c \in B_q \mid |\gamma_2|_c \geq q/2\}$. Let $z = \max\{|X|, |Y|\}$. Then $z \geq |B_q|/2$. We can assume that $z = |X|$, the case $z = |Y|$ is analogical. We also define $C_i = \{c \in X \mid |\gamma_1|_c = q - i\}$ for each $i \in \{1, \ldots, \lfloor q/2 \rfloor\}$. Then we get

$$\left| \bigcup_{i=1}^{\lfloor \frac{q}{2} \rfloor} C_i \right| = z \geq \frac{|B_q|}{2} \geq \sum_{i=1}^{\lfloor \frac{q}{2} \rfloor} [N(N(m,i), q-i) - 1] + 1 \ .$$

Therefore there exists $k \in \{1, \ldots, \lfloor q/2 \rfloor\}$ such that $|C_k| \geq N(N(m,k), q-k)$. Otherwise we would have

$$\left| \bigcup_{i=1}^{\lfloor \frac{q}{2} \rfloor} C_i \right| = \sum_{i=1}^{\lfloor \frac{q}{2} \rfloor} |C_i| \leq \sum_{i=1}^{\lfloor \frac{q}{2} \rfloor} [N(N(m,i), q-i) - 1] \ .$$

Let $\delta_1 = \pi_{C_k}(\gamma_1)$. Then there exists $A' \subseteq \mathrm{alph}(\delta_1)$, $|A'| \geq N(m,k)$, positive integer $p' \in \{1, 2, \ldots, q-k\}$ and words $\delta_{1,1}, \delta_{1,2}, \ldots, \delta_{1,p'}$ such that $\delta_1 = \delta_{1,1} \cdots \delta_{1,p'}$ and $(\delta_{1,j})_{A'}$ is a permutation of A' for all $j \in \{1, \ldots, p'\}$. Let further $\varepsilon_2 = \pi_{A'}(\gamma_2)$. Since $|\mathrm{alph}(\varepsilon_2)| \geq N(m,k)$ and for each $d \in A'$ we have $|\varepsilon_2|_d \leq k$, there exists $A \subseteq \mathrm{alph}(\varepsilon_2)$, $|A| \geq m$, positive integer $p'' \in \{1, 2, \ldots, k\}$ and words $\varepsilon_{2,1}, \varepsilon_{2,2}, \ldots, \varepsilon_{2,p''}$ such that $\varepsilon_2 = \varepsilon_{2,1} \cdots \varepsilon_{2,p''}$ and $(\varepsilon_{2,j})_A$ is a permutation of A for all $j \in \{1, \ldots, p''\}$. If we denote $\varepsilon_1 = \pi_A(\delta_1)$ and $\varepsilon_{1,l} = \pi_A(\delta_{1,l})$ for $l \in \{1, \ldots, p'\}$,

we get $\varepsilon_1 = \varepsilon_{1,1} \cdots \varepsilon_{1,p'}$ and each of the words $(\varepsilon_{1,l})_A$ is a permutation of A. All in all, we have

$$\pi_A(\alpha) = \varepsilon = \varepsilon_1\varepsilon_2 = \varepsilon_{1,1} \cdots \varepsilon_{1,p'}\varepsilon_{2,1} \cdots \varepsilon_{2,p''},$$

where $p' + p'' \leq (q-k)+k = q$. We can factorise the word ε from which it follows that we can also factorise the word α. □

3.2 Estimate of the upper bound

Theorem 4 gives us an upper bound on $N(m,q)$, but the recursive construction is rather complicated and, for practical purposes, difficult to evaluate for large q. Therefore we need an estimate of this bound. Hence we formulate the following theorem.

Theorem 5. *Let $c = 2/3$ and $k = 9/20$. Then $N(m,q) \leq m^{2^{cq}-k}$ for all positive integers m and q.*

Proof. We start with the case $q = 1$. We want

$$N(m,1) = m \leq m^{2^{\frac{2}{3}} - \frac{9}{20}} \approx m^{1.14},$$

which obviously holds for all positive m. Similarly, for $q = 2$ we want

$$N(m,2) = m^2 - m + 1 \leq m^{2^{\frac{4}{3}} - \frac{9}{20}} \approx m^{2.07},$$

which also holds for all positive integers m. Another trivial case is $m = 1$, there we have

$$N(1,q) = 1 = 1^{2^{\frac{2}{3}q} - \frac{9}{20}} \ .$$

For $m = 2$ we need to prove

$$N(2,q) = q + 1 \leq 2^{2^{\frac{2}{3}q} - \frac{9}{20}} \ .$$

We already know, that it holds for $q = 1$ and $q = 2$. Since $2^{\frac{2}{3}q} - 9/20 \geq 2^{\frac{2}{3}q} - 1 \geq q$ for $q \geq 3$ and $2^q \geq q + 1$ for $q \geq 1$, it also holds for $q \geq 3$.

Now we shall focus on the case $q = 3$. For $m = 1$ and $m = 2$ it has been already solved. For other values of m, we shall prove

$$N(m,3) \leq m^{\frac{7}{2}} \leq m^{\frac{71}{20}} = m^{2^{cq}-k} \ .$$

We get

$$N(3,3) \leq 43 \leq 3^{\frac{7}{2}} \approx 46.8 \ .$$

For $m \geq 4$ we have $N(m,3) \leq 2m^3 - m^2 + m$ and thus it remains to be proved that $m^{7/2} \geq 2m^3 - m^2 + m$ for $m \geq 4$. Obviously, $2m^3 \geq 2m^3 - m^2 + m$ and $m^{7/2} = m^3 \cdot \sqrt{m} \geq 2m^3$ for $m \geq 4$.

We proceed by induction on q for $q \geq 4$ and $m \geq 3$. We have

$$N(m,q) \leq m \cdot \left(2 \cdot \left(\sum_{i=1}^{\lfloor \frac{q}{2} \rfloor} [N(N(m,i), q-i) - 1] \right) + N(m, q-1) \right).$$

The induction hypothesis yields

$$\begin{aligned} N(m,q) &\leq m\cdot\left(2\cdot\left(\sum_{i=1}^{\lfloor\frac{q}{2}\rfloor}\left[N(m,i)^{2^{c(q-i)}-k}-1\right]\right)+N(m,q-1)\right)\leq \\ &\leq m\cdot\left(2\cdot\left(\sum_{i=1}^{\lfloor\frac{q}{2}\rfloor}\left[\left(m^{2^{ci}-k}\right)^{2^{c(q-i)}-k}-1\right]\right)+N(m,q-1)\right)= \\ &= m\cdot\left(2\cdot\left(\sum_{i=1}^{\lfloor\frac{q}{2}\rfloor}\left[m^{2^{cq}-k\left(2^{ci}+2^{c(q-i)}\right)+k^2}-1\right]\right)+N(m,q-1)\right)\leq \\ &\leq m\cdot\left(q\cdot M+N(m,q-1)\right), \end{aligned}$$

where

$$M=\max_{1\leq i\leq\lfloor\frac{q}{2}\rfloor}\left\{m^{2^{cq}-k\left(2^{ci}+2^{c(q-i)}\right)+k^2}\right\}\ .$$

It is quite straightforward to see that

$$M=m^{2^{cq}-k\left(2^{c\lfloor\frac{q}{2}\rfloor}+2^{c\lceil\frac{q}{2}\rceil}\right)+k^2}\ .$$

Note that $N(m,q-1)=N(N(m,1),q-1)\leq M$, so we can write $N(m,q)\leq m\cdot(q+1)\cdot M$.

We need to prove that

$$m\cdot(q+1)\cdot m^{-k\left(2^{c\lfloor\frac{q}{2}\rfloor}+2^{c\lceil\frac{q}{2}\rceil}\right)+k^2}\leq m^{-k}$$

for $m\geq 3$ and $q\geq 4$. Denote

$$P(m,q)=\frac{m^{k\left(2^{c\lfloor\frac{q}{2}\rfloor}+2^{c\lceil\frac{q}{2}\rceil}\right)-k^2-k-1}}{q+1}\ .$$

We want to show $P(m,q)\geq 1$ for $m\geq 3$ and $q\geq 4$. Since

$$k\left(2^{c\lfloor\frac{q}{2}\rfloor}+2^{c\lceil\frac{q}{2}\rceil}\right)-k^2-k-1\geq 0$$

for $q\geq 3$, we get $P(m+1,q)\geq P(m,q)$ for $q\geq 3$ and every positive integer m. Furthemore

$$\begin{aligned} \frac{P(m,q+1)}{P(m,q)} &= \frac{(q+1)m^{k\left(2^{c\lceil\frac{q+1}{2}\rceil}+2^{c\lfloor\frac{q+1}{2}\rfloor}-2^{c\lceil\frac{q}{2}\rceil}-2^{c\lfloor\frac{q}{2}\rfloor}\right)}}{q+2}= \\ &= \frac{(q+1)m^{k\left(2^{c\lceil\frac{q+1}{2}\rceil}-2^{c\lfloor\frac{q}{2}\rfloor}\right)}}{q+2}\geq 1 \end{aligned}$$

for $m\geq 3$ and $q\geq 4$, because $(q+1)/(q+2)\geq 5/6$ and

$$m^{k\left(2^{c\lceil\frac{q+1}{2}\rceil}-2^{c\lfloor\frac{q}{2}\rfloor}\right)}\geq 3^{\frac{9}{20}(4-2^{\frac{4}{3}})}\approx 2.08\ .$$

Hence, $P(m, q{+}1) \geq P(m, q)$ for $m \geq 3$ and $q \geq 4$. We have $P(14, 4) \approx 1.015 > 1$, and thus $P(m, q) \geq 1$ for $m \geq 14$ and $q \geq 4$. Similarly $P(5, 5) \approx 1.31 > 1$ and $P(3, 6) \approx 1.21 > 1$, and therefore $P(m, q) \geq 1$ also for $m \geq 5$ and $q \geq 5$ and for $m \geq 3$ and $q \geq 6$. We can verify manually that $N(m, q) \leq m^{2^{cq}-k}$ in the following cases: $q = 4$ and $m \in \{3, \ldots, 13\}$, $q = 5, m = 3$ and $q = 5, m = 4$. Thus, the proof is complete. □

We can now combine the information about $N(m, q)$. By applying the upper bound from [6] and the Theorems 2, 3, 4, and 5 we get that $N(m, q)$ satisfies the following conditions:

$$\begin{aligned}
N(m, 1) &= m, \\
N(m, 2) &= m^2 - m + 1, \\
N(1, q) &= 1, \\
N(2, q) &= q + 1, \\
N(3, 3) &\leq 43, \\
N(m, 3) &\leq 2m^3 - m^2 + m \text{ for } m \geq 4, \\
N(m, q) &\leq m^{2^{cq}}, \text{ where } c = \frac{2}{3} \ .
\end{aligned}$$

4. Multicollision attack on hash functions

In this chapter, we shall describe a multicollision attack on generalised iterated hash functions. Firstly, we introduce basics of collision attacks on (generalised) iterated hash functions and also the general schema of our attack. In the second subchapter, we add a few combinatorial results from [5, 6] and specify the attack.

4.1 General schema of the attack

Suppose a compression function $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^n$ is given. Furthermore, let $l \in \mathbb{N}_+$ and $\tau \in \mathbb{N}_l^+$ be a word such that $\text{alph}(\tau) = \{t\}$, i.e. the alphabet of τ contains only one symbol $t \in \mathbb{N}_l$. Let also $\omega \in \{0,1\}^m$ be a given fixed message block.

We want to find a collision in f_t with the initial value $h \in \{0,1\}^n$. We proceed as follows. Firstly we generate a set $R \subseteq \{0,1\}^m$ of $2^{\frac{n}{2}}$ random message blocks. Then we construct a set S defined by

$$S = \{u_1 u_2 \dots u_l \mid u_t \in R \text{ and } \forall i \in \mathbb{N}_l \setminus \{t\} : u_i = \omega\}.$$

This set contains (at least) two messages with the same hash value with the probability $\tilde{p} \approx 0.4$ (see [7]). By computing $f_t(h,u)$ for each $u \in S$ we can find a collision with the probability $\tilde{p}$. This procedure is a so-called *basic birthday attack*.

If we want to find a collision with probability equal to one, we can repeat the basic birthday attack until a collision is found. It is obvious that the expected number $\tilde{a}$ of basic birthday attacks needed is equal to $1/\tilde{p} \approx 2.5$. This procedure is called an *extended birthday attack*.

Let $H_{\hat{\alpha},f}$ be a generalised iterated hash function. An initial value $h_0 \in \{0,1\}^n$ and $r \in \mathbb{N}_+$ are given and we want to find a 2^r-collision in $H_{\hat{\alpha},f}$ with the initial value h_0. We already know that it can be achieved, with probability approximately $\frac{1}{2}$, by hashing

$$((2^r)!)^{\frac{1}{2^r}} 2^{\frac{n(2^r-1)}{2^r}}$$

messages. Our aim is to find an attack with complexity (expected number of queries) $O(p(n,r)2^{\frac{n}{2}})$, where $p(n,r)$ is a polynomial. This would prove that generalised iterated hash functions are not secure, because multicollisions can be found with less queries than expected of an ideal hash function.

However, the general case, i.e. with no restrictions on $H_{\hat{\alpha},f}$, seems to be very difficult. Therefere, we shall restrict ourselves to a special case of q-bounded hash functions. Let q be a positive integer. We say that the sequence $\hat{\alpha} = (\alpha_1, \alpha_2, \dots)$ is q-bounded if $|\alpha_j|_i \leq q$ for each $j \in \mathbb{N}_+$ and $i \in \mathbb{N}_j$. We call the hash function $H_{\hat{\alpha},f}$ q-bounded, if $\hat{\alpha}$ is q-bounded.

The following schema from [5], called *nested multicollision attack schema*, generally describes our attack.

Input: A generalised iterated hash function $H_{\hat{\alpha},f}$, an initial value $h_0 \in \{0,1\}^n$ and a positive integer r.

Output: A 2^r-collision in $H_{\hat{\alpha},f}$.
Step 1: Choose (a large) $l \in \mathbb{N}_+$. Consider the lth element α_l of the sequence $\hat{\alpha}$. Let $\alpha_l = i_1 i_2 \cdots i_s$, where $s \in \mathbb{N}_+$ and $i_j \in \mathbb{N}_l$ for $j = 1, 2, \ldots, s$.
Step 2: Fix a (large) set of *active indices* $\mathrm{Act} \subseteq \mathbb{N}_l$.
Step 3: Factorise the word α_l into nonempty strings appropriately, i.e. find $p \in \{1, 2, \ldots, s\}$ and $\beta_i \in \mathbb{N}_l^+$ such that $\alpha_l = \beta_1 \beta_2 \cdots \beta_p$.
Step 4: Based upon the active indices, create a large multicollison if f_{β_1}. More exactly, find message block sets $M_1, M_2, \ldots, M_l$ satisfying the following properties.

(i) If $i \in \mathbb{N}_l \setminus \mathrm{Act}$, then the set M_i consists of one constant message block ω.
(ii) If $i \in \mathrm{Act}$, then the set M_i consists of two different message blocks m_{i1} and m_{i2}.
(iii) The set $M = M_1 M_2 \cdots M_l = \{u_1 u_2 \cdots u_l \mid u_i \in M_i, i = 1, 2, \ldots, l\}$ is a $2^{|\mathrm{Act}|}$-collision in f_{β_1} with initial value h_0.

Step 5: Based on the set $C_1 = M$, find message sets $C_2, C_3, \ldots, C_p$ such that

(iv) $C_p \subseteq C_{p-1} \subseteq \cdots \subseteq C_1 = M$.
(v) For each $j \in \{1, 2, \ldots, p\}$ the set C_j is a (large) multicollision in $f_{\beta_1 \beta_2 \cdots \beta_j}$ with initial value h_0.
(vi) $|C_p| = 2^r$.

Step 6: Output C_p.

Evidently, the set C_p forms a 2^r-collision in $H_{\hat{\alpha},f}$. We shall specify the steps in the following subchapter.

4.2 Description of the attack

Before we can describe the attack itself, we need to formulate a few more theorems.

Theorem 6. *Let $k \in \mathbb{N}_+$ and A be a finite nonempty set such that k divides $|A|$. Furthemore, let $\{B_i\}_{i=1}^k$ and $\{C_j\}_{j=1}^k$ be partitions of A such that $|B_i| = |C_j|$ for $i, j = 1, 2, \ldots, k$. Then for each $x \in \mathbb{N}_+$ such that $|A| \geq k^3 \cdot x$, there exists a bijection $\sigma : \{1, 2, \ldots, k\} \to \{1, 2, \ldots, k\}$ for which $|B_i \cap C_{\sigma(i)}| \geq x$ for $i = 1, 2, \ldots, k$.*

For the proof of the previous theorem see Theorem 4.5 in [5].

Theorem 7. *Let $d_0, d_1, \ldots, d_r$, where $r \in \mathbb{N}_+$, be positive integers such that d_i divides d_{i-1} for $i = 1, 2, \ldots, r$, A an alphabet of cardinality $|A| = d_0 d_1^2 d_2^2 \cdots d_r^2$ and $w_1, w_2, \ldots, w_{r+1}$ permutations of A. Then there exists a subset B of A of cardinality $|B| = d_0$ such that the following conditions are satisfied.*

(1) *For any $i \in \{1, 2, \ldots, r\}$, if $\pi_B(w_i) = x_1 x_2 \cdots x_{d_i}$ is the factorisation of $\pi_B(w_i)$ and $\pi_B(w_{i+1}) = y_1 y_2 \cdots y_{d_i}$ is the factorisation of $\pi_B(w_{i+1})$ into d_i equal length $(= \frac{d_0}{d_i})$ blocks, then for each $j \in \{1, 2, \ldots, d_i\}$, there exists $j' \in \{1, 2, \ldots, d_i\}$ such that $\mathrm{alph}(x_j) = \mathrm{alph}(y_{j'})$; and*

(2) *If* $w_{r+1} = u_1u_2\cdots u_{d_r}$ *is the factorisation of* w_{r+1} *into* d_r *equal length* $(= d_0d_1^2d_2^2\cdots d_{r-1}^2d_r)$ *blocks, then* $\pi_B(w_{r+1}) = \pi_B(u_1)\pi_B(u_2)\cdots\pi_B(u_{d_r})$ *is the factorisation of* $\pi_B(w_{r+1})$ *into* d_r *equal length* $(= \frac{d_0}{d_r})$ *blocks.*

For the proof of the previous theorem see Theorem 3 in [6].

Theorem 8. *Let* α *be a word and* $k \geq 2$, $n \geq 1$ *and* $q \geq 2$ *integers such that*

(1) $|\mathrm{alph}(\alpha)| \geq N(n^{(q-1)^2}k^{2q-3}, q)$*; and*
(2) $|\alpha|_a \leq q$ *for each* $a \in \mathrm{alph}(\alpha)$.

Then there exists $B \subseteq \mathrm{alph}(\alpha)$*, an integer* $p \in \{1, 2, \ldots, q\}$ *and a factorisation* $\alpha = \alpha_1\alpha_2\cdots\alpha_p$ *for which*

(3) $|B| = n^{p-1}k$*;*
(4) $B \subseteq \mathrm{alph}(\alpha_i)$ *and the elements of* B *are independent with respect to* $\prec_{\alpha_i}$ *for* $i = 1, 2, \ldots, p$*, i.e. the elements form a chain in* $(\mathrm{alph}(\alpha_i), \prec_{\alpha_i})$*; and*
(5) *for any* $i \in \{1, 2, \ldots, p-1\}$*, if* $(\alpha_i)_B = z_1z_2\cdots z_{n^{p-i}k}$ *is the factorisation of* $(\alpha_i)_B$ *into* $n^{p-i}k$ *equal length* $(= n^{i-1})$ *blocks and* $(\alpha_{i+1})_B = u_1u_2\cdots u_{n^{p-i-1}k}$ *is the factorisation of* $(\alpha_{i+1})_B$ *into* $n^{p-i-1}k$ *equal length* $(= n^i)$ *blocks, then for each* $j_1 \in \{1, 2, \ldots, n^{p-i}k\}$ *there exists* $j_2 \in \{1, 2, \ldots, n^{p-i-1}k\}$ *such that* $\mathrm{alph}(z_{j_1}) \subseteq \mathrm{alph}(u_{j_2})$.

The previous theorem follows directly from Theorems 6 and 7; the proof can also be found in [6]. It is of fundamental importance in our attack schema because it enables us to factorise the word α in the desired way. Now we have the necessary combinatorial background to specify the first three steps of the attack.

Input: A q-bounded ($q \in \mathbb{N}$, $q \geq 2$) generalised iterated hash function $H_{\hat{\alpha},f}$, an initial value $h_0 \in \{0,1\}^n$ and a positive integer r.
Output: A 2^r-collision in $H_{\hat{\alpha},f}$.
Step 1: Let $l = N(n^{(q-1)^2}r^{2q-3}, q)$. Let $\alpha = \alpha_l$ where α_l is the lth element of the sequence $\hat{\alpha}$. Write α in the form $\alpha_l = i_1i_2\cdots i_s$, where $s \in \mathbb{N}_+$ and $i_j \in \mathbb{N}_l$ for $j = 1, 2, \ldots, s$.
Step 2: Let $\mathrm{Act} = B, |B| = n^{p-1}r$, be the set of active indices, where $B \subseteq \mathbb{N}_l$ and $p \in \{1, 2, \ldots, q\}$ are as in Theorem 8, when the parameter $k = r$.
Step 3: Let $\alpha = \beta_1\beta_2\cdots\beta_p$ be the factorisation of α such that the words $\beta_1, \beta_2, \ldots, \beta_p$ have the same properties as the words $\alpha_1, \alpha_2, \ldots, \alpha_p$, respectively, in Theorem 8, when $k = r$.

The following two lemmata are rather technical and we shall not prove them. The proofs can be found in [5]. Remember that $\tilde{a}$ is the expected number of basic birthday attacks needed to find a collision with probability equal to one.

Lemma 2. *Let* α *be a word over the alphabet* $\mathbb{N}_l$, $k \in \mathbb{N}_+$ *and* $a_1, a_2, \ldots, a_k \in \mathrm{alph}(\alpha)$ *symbols such that* $a_1 \prec_\alpha a_2 \prec_\alpha \ldots \prec_\alpha a_k$*. Let furthemore* $\alpha = \alpha_1\alpha_2\cdots\alpha_k$ *be a factorisation of* α *such that for each* $i \in \{1, 2, \ldots, k\}$ *all occurences of the symbol* a_i *in* α *lie in* α_i*. Given an initial value* $h_0 \in \{0,1\}^n$*, we can, with probability equal to one, find message block sets* $M_1, M_2, \ldots, M_l \subseteq \{0,1\}^m$ *as well as values* $h_1, h_2, \ldots, h_k \in \{0,1\}^n$ *such that*

(1) $M_b = \{\omega\}$ *for each* $b \in \mathbb{N}_l \setminus A$, *where* $A = \{a_1, a_2, \ldots, a_k\}$;

(2) $M_{a_i} = \{u_i, u_i'\}$, *where* $u_i \neq u_i'$ *for each* $i \in \{1, 2, \ldots, k\}$;

(3) *for each* $i \in \{1, 2, \ldots, k\}$ *the set* $M = M_1 \cdot M_2 \cdots M_l$ *is a 2-collision in* f_{α_i} *with initial value* h_{i-1} *and a* 2^i*-collision in* $f_{\alpha_1\alpha_2\cdots\alpha_i}$ *such that* $\forall u, u' \in M$

$$\begin{aligned} h_i = f_{\alpha_i}(h_{i-1}, u) &= f_{\alpha_i}(h_{i-1}, u') \quad \textit{and} \\ f_{\alpha_1\alpha_2\cdots\alpha_i}(h_0, u) &= f_{\alpha_1\alpha_2\cdots\alpha_i}(h_0, u'). \end{aligned}$$

Moreover, the expected number of queries on f *needed to carry out the task is* $\tilde{a}|\alpha|2^{\frac{n}{2}}$.

Lemma 3. *Let* α *be a word over the alphabet* $\mathbb{N}_l$, d *and* s *positive integers,* $A \subseteq \mathrm{alph}(\alpha)$ *of cardinality* $|A| = dns$, *and* $\alpha = \beta_1\beta_2\cdots\beta_{ns}\gamma_1\gamma_2\cdots\gamma_s$ *a factorisation of* α *with the following properties.*

(1) $A \subseteq \mathrm{alph}(\beta) \cap \mathrm{alph}(\gamma)$ *where* $\beta = \beta_1\beta_2\cdots\beta_{ns}$ *and* $\gamma = \gamma_1\gamma_2\cdots\gamma_s$;

(2) $|\mathrm{alph}(\beta_i) \cap A| = d$ *for* $i = 1, 2, \ldots, ns$ *and* $|\mathrm{alph}(\gamma_j) \cap A| = nd$ *for* $j = 1, 2, \ldots, s$; *and*

(3) *for each* $i \in \{1, 2, \ldots, ns\}$ *there exists* $j \in \{1, 2, \ldots, s\}$ *such that* $\mathrm{alph}(\beta_i) \cap A \subseteq \mathrm{alph}(\gamma_j) \cap A$.

Furthermore, let $u_1, u_1', u_2, u_2', \ldots, u_{ns}, u_{ns}' \in \{0,1\}^{ml}$ *be messages and* $h_0, h_1, \ldots, h_{ns} \in \{0,1\}^n$ *be values such that for each* $i \in \{1, 2, \ldots, ns\}$:

(4) $\forall b \in \mathbb{N}_l \setminus A : \bar{u}_i(b) = \bar{u}_i'(b)$

(5) $\bar{u}_i(\beta_i) \neq \bar{u}_i'(\beta_i)$ *and* $h_i = f_{\beta_i}(h_{i-1}, u_i) = f_{\beta_i}(h_{i-1}, u_i')$.

Then the set S *of all messages* $u \in \{0,1\}^{ml}$ *such that for each* $b \in \mathbb{N}_l \setminus A : \bar{u}(b) = \omega$ *and for each* $i \in \{1, 2, \ldots, ns\} : \bar{u}(\beta_i) \in \{\bar{u}_i(\beta_i), \bar{u}_i'(\beta_i)\}$ *is well-defined and satisfies for each* $i \in \{1, 2, \ldots, ns\}$ *and* $u \in S$ *the equality* $h_i = f_{\beta_i}(h_{i-1}, u)$. *Moreover we can, with probability equal to one, find messages* $v_1, v_1', v_2, v_2', \ldots, v_s, v_s' \in S$ *and values* $h_0', h_1', \ldots, h_s'$, $h_0' = h_{ns}$, *such that for each* $j \in \{1, 2, \ldots, s\}$

(6) $\bar{v}_j(\gamma_j) \neq \bar{v}_j'(\gamma_j)$ *and* $h_j' = f_{\gamma_j}(h_{j-1}', v_j) = f_{\gamma_j}(h_{j-1}', v_j')$.

The expected number of queries on f *needed to carry out the task is* $\tilde{a}|\gamma|2^{\frac{n}{2}}$. *Finally, the set* T *of all messages* $v \in \{0,1\}^{ml}$ *such that for each* $b \in \mathbb{N}_l \setminus A : \bar{v}(b) = \omega$ *and for each* $j \in \{1, 2, \ldots, s\} : \bar{v}(\gamma_j) \in \{\bar{v}_j(\gamma_j), \bar{v}_j'(\gamma_j)\}$ *is a well-defined subset of* S *and forms a nontrivial* 2^s*-collision in* f_α *with initial value* h_0.

The following theorem, or rather its proof, describes how to proceed in the Steps 4 and 5 of our attack.

Theorem 9. *Let* α *be a word over the alphabet* $\mathbb{N}_l$, r *and* p *positive integers,* A *a subset of the alphabet* $\mathrm{alph}(\alpha)$ *of cardinality* $|A| = n^{p-1}r$, *and* $\alpha = \alpha_1\alpha_2\cdots\alpha_p$ *a factorisation of* α *such that for each* $i \in \{1, 2, \ldots, p\}$, *the elements of* A *form a chain in the partially ordered set* $(\mathrm{alph}(\alpha), \prec_\alpha)$ (*i.e. the elements of* A *are independent with respect to* $\prec_\alpha$). *Assume furthemore that for each* $i \in \{1, 2, \ldots, p\}$, *there exists a factorisation* $\alpha_i = \alpha_{i1}\alpha_{i2}\cdots\alpha_{i,n^{p-i}r}$ *of the word* α_i *such that the following conditions are satisfied.*

(1) $|\mathrm{alph}(\alpha_{ij}) \cap A| = n^{i-1}$ *for each* $i \in \{1, 2, \ldots, p\}$ *and* $j \in \{1, 2, \ldots, n^{p-i}r\}$; *and*

(2) *for every* $i \in \{1, 2, \ldots, p\}$ *and every* $j \in \{1, 2, \ldots, n^{p-i}r\}$ *there exists* $k \in \{1, 2, \ldots, n^{p-i-1}r\}$ *such that* $\mathrm{alph}(\alpha_{ij}) \cap A$ *is a subset of* $\mathrm{alph}(\alpha_{i+1,k}) \cap A$.

Then, given an initial value $h_0 \in \{0,1\}^n$ *we can, with probability equal to one, find a nontrivial* 2^r*-collision in* f_α. *Moreover, the expected number of queries on* f_α *needed to carry out the task is* $\tilde{a}|\alpha|2^{\frac{n}{2}}$.

Remark 5. The previous theorem follows from Lemmata 2 and 3. The complete proof can be found in [5]. Here, we only mention the basic steps of the construction of the multicollision. We use Lemma 2 to create a $2^{n^{p-1}r}$-collision in f_{α_1}, which we then extend, by using Lemma 3 repeatedly, to a 2^r-collision in f_α. Firstly, choose the parameters in Lemma 2 as follows: α is equal to α_1 and k is equal to $n^{p-1}r$. Set $A = \{a_1, a_2, \ldots a_{n^{p-1}r}\}$, where $a_1 \prec_{\alpha_1} a_2 \prec_{\alpha_1} \cdots \prec_{\alpha_1} a_{n^{p-1}r}$. Then Lemma 2 ensures that we can find, with probability equal to one, a $2^{n^{p-1}r}$-collision set M in f_{α_1}. Moreover, the expected number of queries on f is $\tilde{a}|\alpha_1|2^{\frac{n}{2}}$.

Denote $B_1 = M$. Choose the parameters in Lemma 3 as follows: Let β be α_1, γ be α_2 and s be $n^{p-2}r$. Let d be equal 1, β_i be equal to α_{1i} for each $i \in \{1, 2, \ldots, n^{p-1}r\}$ and γ_j be equal to α_{2j} for each $j \in \{1, 2, \ldots, n^{p-2}r\}$. Then the set S in Lemma 3 is equal to B_1. According to the lemma, there exists a subset $T \subseteq S$ that forms a nontrivial $2^{n^{p-2}r}$-collision in $f_{\alpha_1\alpha_2}$ with the initial value h_0. The expected number of queries on f is $\tilde{a}|\alpha_2|2^{\frac{n}{2}}$. Denote $B_2 = T$.

Similarly, let $k \in \{2, 3, \ldots, p-1\}$ and suppose we have a set B_k, which is a $2^{n^{p-k}r}$-collision in $f_{\alpha_1\alpha_2\cdots\alpha_k}$ with the initial value h_0. We apply Lemma 3 with the parameters set as follows: let β be α_k, γ be α_{k+1} and s be $n^{p-k-1}r$. Let d be equal n^{k-1}, β_i be equal to α_{ki} for each $i \in \{1, 2, \ldots, n^{p-k}r\}$ and γ_j be equal to $\alpha_{k+1,j}$ for each $j \in \{1, 2, \ldots, n^{p-k-1}r\}$. Then the set S is equal to B_k and we can find $T \subseteq S = B_k$ forming a $2^{n^{p-k-1}r}$-collision in $f_{\alpha_1\alpha_2\cdots\alpha_k}$ with the initial value h_0. Denote $B_{k+1} = T$. By repeating this process we can find a 2^r-collision in f_α with the initial value h_0. The expected number of queries on f is equal to $\tilde{a}|\alpha|2^{\frac{n}{2}}$.

Therefore we can define the last three steps of the attack as follows:

Step 4: Let $M_1, M_2, \ldots, M_l$ be as in Lemma 2 where we set $A = \mathrm{Act}$ and α is equal to β_1 from Step 3.
Step 5: For each $i \in \{1, 2, \ldots, p\}$ let C_i be equal to B_i from the proof of Theorem 9 (see Remark 5), where α is equal to $\alpha = \alpha_l$, $A = \mathrm{Act}$ and p is equal to p from Step 2.
Step 6: Output C_p.

The following theorem sums up our results.

Theorem 10. *Let* m, n *and* q *be positive integers such that* $m > n$ *and* $q \geq 2$, f *a compression function of block size* m *and length* n, *and* $\hat{\alpha} = (\alpha_1, \alpha_2, \ldots)$ *a* q*-bounded sequence of words such that* $\mathrm{alph}(\alpha_i) = \mathbb{N}_i$ *for each* $i \in \mathbb{N}_+$. *Then, for each* $k \in \mathbb{N}_+$, *there exists a* 2^k*-collision attack on the generalised iterated hash function* $H_{\hat{\alpha},f}$ *such that the expected number of queries on* f *is not greater than*

$$\tilde{a} \cdot q \cdot N(n^{(q-1)^2}k^{2q-3}, q)2^{\frac{n}{2}}.$$

Proof. We use our attack schema. In Step 1, we choose $\alpha = \alpha_l$, where $l = N(n^{(q-1)^2}k^{2q-3}, q)$. Theorem 9 implies that the complexity of the attack is $\tilde{a}|\alpha_l|2^{\frac{n}{2}}$. Since α is q-bounded, $|\alpha_l| \leq q \cdot N(n^{(q-1)^2}k^{2q-3}, q)$. □

Conclusion

We have found a lower upper bound on $N(m, q)$; thus we have reduced the complexity of the attack in Theorem 10. These results show that even q-bounded generalised iterated hash functions are not secure, because we can still find multicollisions in them with the complexity lower than expected from an ideal hash function. The problem of generalised iterated hash functions that are not q-bounded is still open.

We should also mention that there does not exist any practical (i.e. not only theoretical) hash function based on the generalised iterated structure. There are several reasons for that. First of all, this structure is computationally demanding, and we usually require hash functions to be fast. Another reason is that the message must be complete before we can start hashing. In the traditional iterated structure, we can start hashing as soon as we have the first block of the message, but in the generalised iterated structure, we need the whole message to determine the number of blocks. To avoid this problem we can fix the number of blocks, and then start writing the message. However, in this case the blocks occuring at the end of the message are not available at the beginning of the hashing. Thus, there are some restrictions on the sequence $\hat{\alpha}$, which can potentially result in reduced security. Last but not least, there arises the problem how to efficiently encode the (infinite) sequence $\hat{\alpha}$. When the sequence is q-bounded, the structure is not secure. When the sequence is not q-bounded, it must have some pattern, which can also lead to attacks.

The upper bound on $N(m, q)$ presented in Chapter 3 is still probably not optimal for most of the cases and could be lowered. This would result in even smaller complexity of the attack. Therefore, the behaviour of $N(m, q)$ could be the object of further study.

Bibliography

[1] DAMGÅRD, I. B. A design principle for hash functions. In *Advances in Cryptology - CRYPTO '89*. New York: Springer, 1990, pp. 416–427. ISBN 978-0-387-97317-3. (Lecture Notes in Computer Science; vol. 435.)

[2] DILWORTH, R. P. A decomposition theorem for partially ordered sets. *Annals of Mathematics*, 1950, **51**(1), 161–166. ISSN 0003486X.

[3] HOCH, J. – SHAMIR, A. Breaking the ICE - finding multicollisions in iterated concatenated and expanded (ICE) hash functions. In *Fast Software Encryption, 13th International Workshop, FSE 2006, Graz, Austria, March 15-17, 2006, Revised Selected Papers*. Springer, 2006, pp. 179–194. ISBN 978-3-540-36597-6. (Lecture Notes in Computer Science; vol. 4047).

[4] JOUX, A. Multicollisions in iterated hash functions. Application to Cascaded Constructions. In *Advances in Cryptology - CRYPTO 2004, 24th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings*. Springer, 2004, pp. 303–316. ISBN 978-3-540-22668-0. (Lecture Notes in Computer Science; vol. 3152).

[5] KORTELAINEN, J. – HALUNEN, K. – KORTELAINEN, T. Multicollision attacks and generalized iterated hash functions. *Journal of Mathematical Cryptology*, 2011, **4**(3), 239–270. ISSN 1862-2976.

[6] KORTELAINEN, J. – KORTELAINEN, T. – VESANEN, A. Unavoidable regularities in long words with bounded number of symbol occurrences. In *Computing and Combinatorics, 17th Annual International Conference, COCOON 2011, Dallas, TX, USA, August 14-16, 2011. Proceedings*. Berlin: Springer, 2011, pp. 519–530. ISBN 978-3-642-22684-7. (Lecture Notes in Computer Science; vol. 6842).

[7] MENEZES, A. J. – VAN OORSCHOT, P. C. – VANSTONE, S. A. *Handbook of Applied Cryptography*. CRC Press, 1996. ISBN 978-0849385230.

[8] MERKLE, R. C. One Way Hash Functions and DES. In *Advances in Cryptology - CRYPTO '89*. New York: Springer, 1990, pp. 428–446. ISBN 978-0-387-97317-3. (Lecture Notes in Computer Science; vol. 435.)

[9] NANDI, M. – STINSON, D. Multicollision attacks on some generalized sequential hash functions. *IEEE Transactions on Information Theory*, 2007, **53**(2), 759–767. ISSN 0018-9448.

[10] SUZUKI, K. – TONIEN, D. – KUROSAWA, K. – TOYOTA, K. Birthday paradox for multi-collisions. In *Information Security and Cryptology – ICISC 2006, 9th International Conference, Busan, Korea, November 30 - December 1, 2006, Proceedings*. Berlin: Springer, 2006, pp. 29–40. ISBN 978-3-540-49112-5. (Lecture Notes in Computer Science; vol. 4296).